



Using ArrayOL to Identify Potentially Shareable Data in Thread Work-Groups of GPUs

Antonio Wendell de Oliveira Rodrigues, Frédéric Guyomarc'H, Jean-Luc Dekeyser

► To cite this version:

Antonio Wendell de Oliveira Rodrigues, Frédéric Guyomarc'H, Jean-Luc Dekeyser. Using ArrayOL to Identify Potentially Shareable Data in Thread Work-Groups of GPUs. Designing for Embedded Parallel Computing Platforms: Architectures, Design Tools, and Applications on DATE 2011, Mar 2011, Grenoble, France. inria-00594304

HAL Id: inria-00594304

<https://inria.hal.science/inria-00594304>

Submitted on 19 May 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Using ArrayOL to Identify Potentially Shareable Data in Thread Work-Groups of GPUs

Wendell Rodrigues, Frédéric Guyomarc'h, Jean-Luc Dekeyser

LIFL - USTL

INRIA Lille Nord Europe - 59650

Villeneuve d'Ascq - France

{wendell.rodrigues, frederic.guyomarch, jean-luc.dekeyser}@inria.fr

Abstract—Over recent years, using Graphics Processing Units (GPUs) has become as an effective method for increasing the performance of many applications. However, these performance benefits from GPUs come at a price. First, extensive programming expertise and intimate knowledge of the underlying hardware are essential for gaining good speedups. Second, the expressibility of GPU-based programs are not powerful enough to retain the high-level abstractions of the solutions. Although the programming experience has been significantly improved by existing frameworks like CUDA and OpenCL, it is still a challenge to effectively utilise these devices while still retaining the programming abstractions. To this end, performing a model-to-source transformation, whereby a highlevel language is mapped to CUDA or OpenCL, is an attractive option. In particular, it enables to harness the power of GPUs without any expertise on the GPGPU programming. In this work, we propose an approach based on MDE and ArrayOL to detect shareable data zone. The tilers from ArrayOL, which allow express the data parallelism from repetitive tasks, are analyzed in time compilation to create areas of shared data. The identification of these areas is crucial to allow us loading data on shared areas of memory that have high throughput. Consequently, programs automatically generated shall have performances comparable to manually well written programs.

Index Terms—GPU, Embedded Systems, MDE